

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
 - TEXT CUT OFF AT TOP, BOTTOM OR SIDES
 - FADED TEXT
 - ILLEGIBLE TEXT
 - SKEWED/SLANTED IMAGES
 - COLORED PHOTOS
 - BLACK OR VERY BLACK AND WHITE DARK PHOTOS
-
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **10312383 A**(43) Date of publication of application: **24 . 11 . 98**

(51) Int. Cl.

G06F 17/28(21) Application number: **10008593**(22) Date of filing: **20 . 01 . 98**(30) Priority: **11 . 03 . 97 JP 09 56301**(71) Applicant: **TOSHIBA CORP**(72) Inventor: **ITO ETSUO**

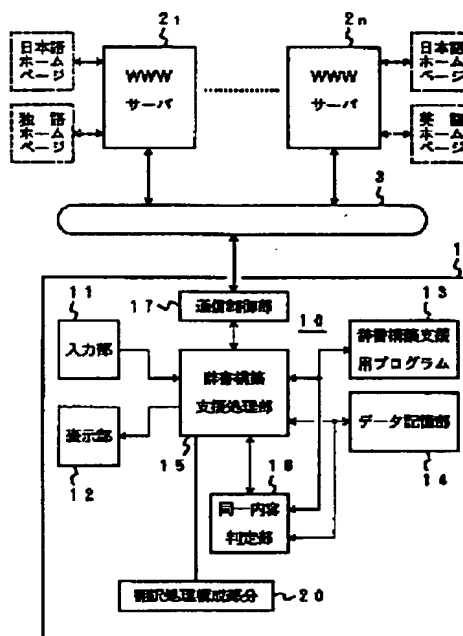
(54) **TRANSLATION SYSTEM, TRANSLATION
PROCESSING METHOD, AND RECORDING
MEDIUM FOR TRANSLATION PROCESSING**

(57) Abstract:

PROBLEM TO BE SOLVED: To obtain translation data in quantities.

SOLUTION: This dictionary structuring support system is provided with a document input means 15 which specifies ULR's on a document screen of a hypertext and takes in documents regarding homepages of different languages from WWW servers, an identical content decision part 16 which compares data on image data, graphs, patterns, link destination, etc., of the homepages of both the languages taken in by the document input means and decides whether or not they have the same meaning and contents, and a means 15 which extracts a pair of translation data of the original text and its translation from the documents regarding the homepages of both the languages when identity is decided and registers them in a translation dictionary.

COPYRIGHT: (C)1998,JPO



Our Ref: OP1081-US

Prior Art Reference:

Japanese Patent Laid-Open Publication No. 10-312383
Date of Publication: November 24, 1998
Title: TRANSLATION SYSTEM, METHOD OF TRANSLATION PROCESS,
AND RECORDING MEDIUM FOR TRANSLATION PROCESS
Patent Application No. 10-8593
Filing Date: January 20, 1998
Internal Priority Application No. 9-56301
Priority Date: March 11, 1997
Country: Japan
Applicant: ID No. 000003078
KABUSHIKI KAISHA TOSHIBA (TOSHIBA CORP.)
Kawasaki-shi, Kanagawa-ken, Japan
Inventor: Etsuo ITOH
c/o Toshiba Fuchu Plant
Fuchu-shi, Tokyo, Japan

Partial Translation

[ABSTRACT]

[Object] To enable to obtain a translation data in large quantities.

[Solving Means] A dictionary editing support system comprising document take-in means 15 which designates a ULR of a hyper text document screen and takes-in documents relating to home pages in different languages from a WWW server; content identicalness determination unit 16 which compares data, such as, image data, graphs, patterns and destinations of linkage which are present in the home page documents in two languages which have been taken-in by the document take-in means, and determines whether these documents have the same meaning and content; and wherein means 15, which, when determines that there is the identicalness, takes out a translation data in a pair of the original language text and a translated language thereof from the documents of home

pages in two languages, and registers the taken-out translation data in a translation dictionary.

Pertinent Description

[0034]

The system of the present invention is achieved in view of and based on the nature of home pages on the Internet. The operation of the system will be described below. Based on the instruction from the input unit 11, the initial screen or the required screen is displayed on the display unit 12, whereupon a destination on the screen (ULR) is clicked, and the dictionary editing support process unit 15 executes the dictionary editing support process by reading a program stored in the recording medium 13 for dictionary editing support process.

[0035]

The dictionary editing support process transmits a ULR address which is based on the instruction from the input unit 11 and a home page taking-in command to, for example, the WWW server 2₁, via the communication management and control unit 17, and takes in the document of, for example, an English language home page from the WWW server 2₁ and stores it in the data storage unit 14, while displays it on the display unit 12.

[0036]

When this English language home page describes, for example, "here for Japanese language", the indicated portion is clicked and the Japanese language home page is also taken-in, which is registered in the data storage unit 14, while being displayed on the display unit. Inversely, when the document in the Japanese language home page is taken-in, and it contains the description of "for English", that indicated portion is clicked similarly, and the document of the English language home page is taken-in.

[0037]

Thereafter, the dictionary editing support process unit 15 requests the content identicalness determination unit 16 to determine whether the meaning and content of the documents of home pages in two languages. The content identicalness determination unit 16 determines whether the image data, graphs, patterns and further the destination of linkage present in the documents of the two home pages are identical or not. This is because, in many a case, when the same image data, graphs, and the like are used, the meaning and content of home pages in two languages are same.

[0038]

Further, according to the necessity, and on the condition that the data of image data, graphs, patterns and further the destination of linkage are identical, or even not on the condition of identicalness of these data of image data, graphs, patterns and the designation of linkage, the determination unit 16 compares the texts described in the home pages in two languages based on the grammar of each language, and determines a percentage(%) of identicalness. Here, if identicalness is at least 80% or more, the remaining 20% or less of the texts is determined as identical, and the determination unit 16 informs the dictionary editing support process unit 15 of such result.

[0039]

Here, the dictionary editing support process unit 15 takes out an original word and its translation in pair successively from the documents of home pages in Japanese and English languages, and registers them in the paired form in the data storage unit 14 or in a translation dictionary of an unillustrated translation process configuration unit 20.

[0040] (up to line 22)

Thus, according to the above-described embodiment, documents of home pages in two different languages from the same WWW server

are taken-in, the identicalness of the two documents is checked under a certain condition, and the identicalness is determined. Thus, the texts described in the home pages in two languages can be obtained as a pair of original language and its translation. Considering the fact that there exists a tremendous number of home pages transmitted internally and from foreign countries, a large quantity of a pair of original language phrase and its translation can be obtained. Further, with this system, the dictionary editing support process is executed automatically based on the dictionary editing support program, the pair of original language and its translation can be obtained without involving time and cost to a great extent. In addition, a precise translation dictionary can be easily configured.

//////////////// The End //////////////////

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-312383

(43) 公開日 平成10年(1998)11月24日

(51) Int. Cl. ⁶

G 0 6 F 17/28

識別記号

F I

G 0 6 F 15/38

C

審査請求 未請求 請求項の数 8

O L

(全12頁)

(21) 出願番号 特願平10-8593

(22) 出願日 平成10年(1998)1月20日

(31) 優先権主張番号 特願平9-56301

(32) 優先日 平9(1997)3月11日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 伊藤 悦雄

東京都府中市東芝町1番地 株式会社東芝
府中工場内

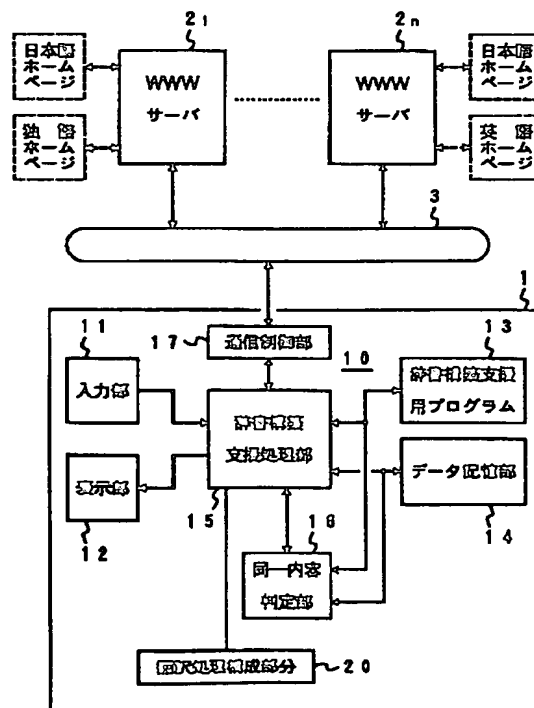
(74) 代理人 弁理士 鈴江 武彦 (外6名)

(54) 【発明の名称】 翻訳システム、翻訳処理方法および翻訳処理用記録媒体

(57) 【要約】

【課題】 訳文データを大量に入手可能とすることにある。

【解決手段】 ハイパーテキストの文書画面のU L Rを指定し、WWWサーバから異なる言語のホームページに係わる文書を取り込む文書取り込み手段15と、この文書取り込み手段で取り込んだ両言語のホームページの文書に存在するイメージデータ、グラフ、図柄、リンク先などのデータを比較し、同一の意味、内容を有するか否かを判定する同一内容判定部16と、同一性有りと判定されたとき、両言語のホームページに係わる文書から原文と訳語との対をなす訳文データを取り出して翻訳辞書に登録する手段15とを設けた辞書構築支援システム。



【特許請求の範囲】

【請求項1】 WWWサーバから異なる言語で同一の意味、内容を有するホームページに係わる文書を取り込んで翻訳辞書用対訳データとして入手する機能を設けたことを特徴とする辞書構築支援システム。

【請求項2】 ハイパーテキストの文書画面の行き先（URL）を指定し、WWWサーバから第1言語のホームページに係わる文書を取り込むとともに、当該第1言語のホームページに係わる文書のリンク先から第1言語以外の第2言語のホームページに係わる文書を取り込む文書取り込み手段と、この文書取り込み手段で取り込んだ両言語のホームページに係わる文書の同一性を判定する同一内容判定手段と、この判定手段によって同一性有りと判定されたとき、両言語のホームページに係わる文書から原文と訳語との対をなす訳文データを取り出して翻訳辞書に登録する手段とを備えたことを特徴とする辞書構築支援システム。

【請求項3】 同一内容判定手段は、両言語のホームページに係わる文書に存在するイメージデータ、グラフ、図柄、リンク先などのデータを比較し、同一の意味、内容を有するか否かを判定することを特徴とする請求項2に記載の辞書構築支援システム。

【請求項4】 WWWサーバから異なる言語のホームページに係わる文書を取り込む文書取り込みステップと、この文書取り込みステップで取り込んだ両言語のホームページに係わる文書の同一性を判定する同一性判定ステップとを有し、同一性有りと判定されたとき、両言語のホームページに係わる文書から翻訳辞書構築用対訳データを入手することを特徴とする辞書構築支援方法。

【請求項5】 対訳データから翻訳辞書を構築するプログラムを読み込み可能に記録した辞書構築支援処理用記録媒体であって、辞書構築支援システムのコンピュータに、

前記ハイパーテキストの文書画面の行き先（URL）を指定し、WWWサーバから第1言語のホームページに係わる文書を取り込むとともに、当該第1言語のホームページに係わる文書のリンク先から第1言語以外の第2言語のホームページに係わる文書を取り込む文書取り込み機能と、

この文書取り込み機能で取り込んだ両ホームページに係わる文書の同一性を判定する同一性判定機能と、

この同一性判定機能によって同一性有りと判定されたとき、両言語のホームページに係わる文書から原文と訳語との対をなす訳文データを取り出し翻訳辞書に登録する辞書構築データ登録機能とを実現させるためのプログラムを記録したコンピュータ読み取り可能な辞書構築支援処理用記録媒体。

【請求項6】 ハイパーテキストの文書画面の行き先（URL）を指定し、WWWサーバから原文となる言語のホームページに係わる文書または当該文書のリンク先

の文書を取り込む原文文書取り込み手段と、この原文文書取り込み手段によって取り込んだ原文となる言語のホームページに係わる文書または当該文書のリンク先の文書を翻訳する際、複数の訳語が存在するとき、前記原文となる言語のホームページに係わる文書のリンク先に対訳言語の文書が存在するとき、当該対訳言語の文書を参照し、その中に原文の単語に対応する訳語があれば、当該訳語を優先選択して翻訳する訳語選択参照手段とを備えたことを特徴とする翻訳システム。

10 【請求項7】 WWWサーバから原文となる言語のホームページに係わる文書または当該文書のリンク先の文書を翻訳するとき、前記原文となる言語のホームページに係わる文書のリンク先の対訳言語の文書を参照し、原文の単語に対応する訳語を選択し翻訳することを特徴とする翻訳処理方法。

【請求項8】 WWWサーバから原文となる言語のホームページに係わる文書または当該文書のリンク先の文書を翻訳するプログラムを読み込み可能に記録した翻訳処理用記録媒体であって、翻訳システムのコンピュータ

20 に、
前記ハイパーテキストの文書画面の行き先（URL）を指定し、WWWサーバから原文となるホームページに係わる文書または当該文書のリンク先の文書を取り込む文書取り込み機能と、

この文書取り込み機能によって取り込んだ原文となるホームページに係わる文書または当該文書のリンク先の文書を翻訳辞書に登録する訳語を用いて翻訳処理する翻訳処理機能と、

この翻訳処理機能によって前記文書の翻訳をする際、複数の訳語が存在するとき、前記原文となるホームページに係わる文書のリンク先に対訳言語の文書が存在するとき、当該対訳言語の文書を参照し、その中に原文の単語に対応する訳語があれば、当該訳語を優先選択して翻訳する訳語選択参照機能と、

を実現させるためのプログラムを記録したコンピュータ読み取り可能な翻訳処理用記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、辞書を構築するための辞書構築支援システム、機械翻訳システム、翻訳処理方法および翻訳処理用記録媒体に関する。

【0002】

【従来の技術】機械翻訳システムは、原文を翻訳処理するときに用いる訳語を記憶する翻訳辞書部およびある言語の原文に対して翻訳辞書から複数の訳語を取り出し、その中から適切と思われる1つの訳語を選択し所要の言語に翻訳する翻訳処理部などで構成されている。

【0003】ところで、翻訳辞書の構築にあっては、専門分野の辞書（書籍）から原文の各単語ごとに適切と思われる複数の訳語を取り出して翻訳辞書に登録すると

か、或いは予め既に多数の対訳データが用意されていれば、その対訳データを用いて翻訳辞書を構築することは既に知られている。

【0004】

【発明が解決しようとする課題】しかしながら、以上のような翻訳辞書を構築するに際し、各原文の単語ごとに同一の意味、内容をもつ対訳データが大量に必要となるが、現在のところ実際の翻訳処理の実用に耐えうるだけの対訳データの入手が困難である。

【0005】また、翻訳辞書には各原文の単語ごとに複数の訳語が登録されているので、ある1つの原文の翻訳に対し、複数の訳語が選択対象として出てくることが多い。今、1つの原文である例えば「PC」に対して、日本語の訳語では例えば「パソコン」、「プログラム カウンター」、「プログラマブル コントローラ」、「プリント回路」、「加圧コンクリート」などが出てくるが、これらの中から何れを用いるかを判断する必要がある。従来の機械翻訳システムでは、文書の分野、他の語句との関係から適切な訳語を判断するが、例えば「girl」を「少女」と訳すか、「乙女」と訳すかはそのどちらでも判断できなかった。

【0006】さらに、翻訳辞書の構築にあつては、翻訳の品質を高める観点から、原文文書と訳文文書との対が必要となるが、これら対となる対応文書を得ることが難しい。

【0007】本発明は上記実情に鑑みてなされたもので、WWWサーバから辞書構築の対応文書を大量に取得可能とする辞書構築支援システムを提供することにある。また、本発明は、WWWサーバから辞書構築用の対応文書を大量に取得する辞書構築支援方法を提供することにある。

【0008】また、本発明は、WWWサーバから取り込んだ異なる言語のホームページに係わる文書から辞書構築用の対応文書を取得するための辞書構築支援用記録媒体を提供することにある。

【0009】さらに、本発明は、ある言語を翻訳するに際し、複数の訳語が存在する時、適切な訳語を迅速に選択可能とする翻訳システムを提供することにある。さらに、本発明は、ある言語を翻訳するに際し、複数の訳語が存在する時、適切な訳語を迅速に選択する翻訳処理方法を提供することにある。さらに、本発明は、ある言語を翻訳するに際し、複数の訳語が存在する時、適切な訳語を選択するための翻訳処理用記録媒体を提供することにある。

【0010】

【課題を解決するための手段】上記課題を解決するために、請求項1～3に対応する発明は、ハイパーテキストの文書画面の行き先(ULR)を指定し、WWWサーバから第1言語のホームページに係わる文書を取り込むとともに、当該第1言語のホームページに係わる文書のリ

ンク先から第1言語以外の第2言語のホームページに係わる文書を取り込む文書取り込み手段と、この文書取り込み手段で取り込んだ両言語のホームページの文書に存在するイメージデータ、グラフ、図柄、リンク先などのデータを比較し、同一の意味、内容を有するか否かを判定する同一内容判定手段と、この判定手段によって同一性有り判定されたとき、両ホームページに係わる文書から原文と訳語との対をなす訳文データを取り出して翻訳辞書に登録する手段とを設けた辞書構築支援システムである。

【0011】この請求項1～3に対応する発明は、以上のような手段を講じたことにより、両言語のホームページの文書に存在するイメージデータ、グラフ、図柄、リンク先などのデータが同一性を有するとき、両文書に記載される文章の単語も同じ意味、内容をもって対応しているため、両文書から原文と訳語との対となる対訳データを容易に入手できる。

【0012】請求項4に対応する発明は、WWWサーバから異なる言語のホームページに係わる文書を取り込む文書取り込みステップと、この文書取り込みステップで取り込んだ両言語のホームページの同一性を判定する同一性判定ステップとを有し、同一性有り判定されたとき、両言語のホームページに係わる文章を翻訳辞書構築用対訳データとして入手するので、WWWサーバから辞書構築用の対応文書を大量に取得することができる。

【0013】請求項5に対応する発明は、対訳データから翻訳辞書を構築するプログラムを記録した辞書構築支援処理用記録媒体であつて、辞書構築支援システムのコンピュータに、前記ハイパーテキストの文書画面の行き先(ULR)を指定し、WWWサーバから第1言語のホームページに係わる文書を取り込むとともに、当該第1言語のホームページに係わる文書のリンク先から第1言語以外の第2言語のホームページに係わる文書を取り込む文書取り込み機能と、この文書取り込み機能で取り込んだ両ホームページの同一性を判定する同一性判定機能と、この同一性判定機能によって同一性有り判定されたとき、両言語のホームページに係わる文書から原文と訳語との対をなす訳文データを取り出し翻訳辞書に登録する辞書構築データ登録機能とを実現させるためのプログラムを記録したコンピュータ読み取り可能な辞書構築支援処理用記録媒体である。

【0014】この請求項5に対応する発明は、以上のような手段を講じたことにより、ハイパーテキストの文書画面の行き先を指定すれば、行き先のWWWサーバから異なる言語のホームページに係わる文書を取り込むことができ、また同一性判定機能により、両言語のホームページに係わる文書の中に存在する例えばイメージデータ、グラフ、図柄、リンク先などのデータに同一性を有するとき、両言語のホームページに係わる文書に記載される文章が同一と判断し、原文と訳語とを対の形で取り

出して翻訳辞書の対訳データとして利用することができる。

【0015】請求項6に対応する発明は、ハイパーテキストの文書画面の行き先を指定し、WWWサーバから原文となる言語のホームページに係わる文書または当該文書のリンク先文書を取り込む原文文書取り込み手段と、この原文文書取り込み手段によって取り込んだ原文となる言語のホームページに係わる文書または当該文書のリンク先の文書を翻訳する際、複数の訳語が存在するとき、前記原文となる言語のホームページに係わる文書のリンク先に対訳言語の文書が存在するとき、当該対訳言語の文書を参照し、その中に原文の単語に対応する訳語があれば、当該訳語を優先選択して翻訳する訳語選択参照手段とを設けた翻訳システムである。

【0016】この請求項6に対応する発明は、以上のような手段を講じたことにより、WWWサーバから原文となる言語のホームページに係わる文書を取り込み、翻訳辞書を用いて翻訳するが、このとき文書のある単語に複数の訳語があれば、原文となる言語のホームページに係わる文書のリンク先に訳語の言語をもつ文書の中に前記単語の訳語があれば、それを優先的に選択し翻訳するので、適切な訳語を用いて精度よく翻訳できる。

【0017】請求項7に対応する発明は、WWWサーバから原文となる言語のホームページに係わる文書または当該文書のリンク先文書を翻訳する際、原文となる言語のホームページに係わる文書のリンク先の対訳言語の文書を参照し、原文の単語に対応する訳語を選択し翻訳することにより、適切な訳語を迅速に選択できる。

【0018】請求項8に対応する発明は、WWWサーバから原文となる言語のホームページに係わる文書または当該文書のリンク先の文書を翻訳するプログラムを記録した翻訳処理用記録媒体であって、翻訳システムのコンピュータに、前記ハイパーテキストの文書画面の行き先(ULR)を指定し、WWWサーバから原文となるホームページに係わる文書または当該文書のリンク先の文書を取り込む文書取り込み機能と、この文書取り込み機能によって取り込んだ原文となるホームページに係わる文書または当該文書のリンク先の文書を翻訳辞書に登録する訳語を用いて翻訳処理する翻訳処理機能と、この翻訳処理機能によって前記文書の翻訳をする際、複数の訳語が存在するとき、前記原文となるホームページに係わる文書のリンク先に対訳言語の文書が存在するとき、当該対訳言語の文書を参照し、その中に原文の単語に対応する訳語があれば、当該訳語を優先選択して翻訳する訳語選択参照機能とを実現させるためのプログラムを記録したコンピュータ読み取り可能な翻訳処理用記録媒体である。このような翻訳処理用記憶媒体によれば、ある言語を翻訳するに際し、複数の訳語が存在する時、適切な訳語を選択し翻訳処理できる。

【0019】

【発明の実施の形態】

(辞書構築支援システムに係わる発明の実施の形態) 図1は本発明に係わる辞書構築支援システムの一実施の形態を示す機能ブロック図である。

【0020】このシステムは、WWWブラウザで作成されるハイパーテキストなどの文書をもつコンピュータによって構成される機械翻訳システム1と、同じくコンピュータで構成され、例えば日本語ホームページに係わる文書および日本語ホームページと対の関係にある日本語以外の言語で作成されるホームページに係わる文書をもつ複数のWWWサーバ2₁, ..., 2_nがネットワーク3に接続されている。

【0021】この機械翻訳システム1は、大きく分けて第1の言語を第2の言語に翻訳処理するための訳語を決定し登録する辞書構築支援システム10と、ある言語の原文に対して構築された翻訳辞書の訳語を用いて翻訳処理を行う翻訳処理構成部分20とから成る。

【0022】この辞書構築支援システム10は、入力部11、表示部12、辞書構築支援用記録媒体13、データ記憶部14、CPUで構成される辞書構築支援処理部15、同一内容判定部16および通信制御部17等によって構成されている。

【0023】入力部11は、ハイパーテキストの文書画面表示指示その他必要なデータ、コマンドを入力するもので、通常、キーボード、OCR、タブレット、マウス、ライトペン或いはFDD、ネットワークなどが用いられ、ハイパーテキストの文書画面の絵または文章などの行き先(ULR)をマウスでクリックすることにより、行き先となるWWWサーバの例えば日本語および日本語以外の言語のホームページに係わる文書を取り込む機能をもっている。

【0024】前記表示部12は、ハイパーテキストの文書画面の表示、WWWサーバから取り込むホームページデータの表示その他入力部11から入力されるデータやデータ記憶部14に記憶されるデータ等を表示するものであって、CRT、ビットマップディスプレイ、投影される壁面やゴーグルなどが用いられる。

【0025】前記辞書構築支援用記録媒体13は、例えば後記する図3に示すような辞書構築支援プログラムを格納するためのものであって、通常、磁氣的、電氣的、光学的、機械的などの記憶素子、つまり、FDD、CD-ROM、磁気ディスク、光ディスクなどが用いられる。また、この記憶媒体13は、辞書構築支援プログラムだけでなく、必要に応じて翻訳処理上必要なプログラム例えば画像表示用プログラム、通信制御用プログラムなど記憶する。

【0026】前記データ記憶部14は、WWWサーバから取り込んだ日本語または日本語以外の言語のホームページに係わる文書データや辞書構築支援処理過程のデータ、さらに入力部11から入力されるデータなどを

記憶するものであって、磁氣的、電氣的、光學的、機械的な記憶素子などが用いられる。

【0027】前記辞書構築支援処理部15は、辞書構築支援用記録媒体13に記憶される辞書構築支援用プログラムを用いて辞書構築支援処理を実行したり、入力部11から入力されるホームページデータの取り込みアドレスまたはコマンドなどを出力する機能をもっている。

【0028】前記同一内容判定部16は、WWWサーバから取り込んだ例えば日本語ホームページに係わる文書と日本語以外の言語のホームページに係わる文書とが同一の意味、内容をもっているか否かを判定し、互いに同一の意味、内容をもっているとき、原文文書（例えば英語ホームページ文書）と訳文文書（例えば日本語ホームページ文書）が対応関係にあって対訳データ可能と判断する。この判定処理にあっては、一定の条件を必要とする。それは、両ホームページに係わる文書の中に存在するイメージデータ、グラフ、図柄、さらにリンク先などが同一であるとき、または割合が一定値以上のとき、両言語のホームページに記載される文章が同じ意味、内容を異なる2つの言語で表現していると判定する。さらに、この判定処理は、必ずしもイメージデータ、グラフ、図柄、さらにリンク先その他全部が完全に一致しても、記載されている文章が異なる場合もあるので、さらに文章どうしを比較し、少なくとも例えば80%以上同一であれば、残りの20%も同一であると判定してもよい。

【0029】前記通信制御部17は、リンク先アドレスの発信やホームページの取り込みアドレスまたはコマンドの発信、ホームページの取り込みなどを行う機能をもっている。

【0030】前記機械翻訳処理構成部分20は、ある言語の原文を所要とする他の言語に翻訳処理する従来公知の構成を用いるものであって、例えばある言語の原文を入力したり、他のコンピュータのFDから原文文書を入力したり、或いはネットワークから送られてくるデータを入力する入力部、この入力部から入力されるデータや翻訳結果を表示する表示部、翻訳辞書や翻訳上の文法等を記憶するデータ記憶部、入力部から入力される原文を翻訳辞書や翻訳上の文法を用いて所定の言語に翻訳する翻訳処理部その他必要な構成部分をもっている。

【0031】なお、機械翻訳処理構成部分20の入力部、表示部、データ記憶部は、辞書構築支援システム0の入力部11、表示部12、データ記憶部14等をそのまま用いてもよい。

【0032】次に、以上のような辞書構築支援システム10の動作を説明するに先立ち、WWWサーバのホームページに係わる文書を対訳データとすることの意義を説明する。

【0033】例えばインターネットのWWWサーバ21、…、2nの中には、自国語および外国語向けのホー

ムページを作成し開示している場合があるが、例えば英語で記載されているホームページの文書に「日本語はここ」とか、逆に日本語のホームページの文書に「for English」とか記載されているとき、そのリンク先部分をクリックすれば、日本語と英語のホームページに係わる文書をダウンロードできる。しかし、これら両言語のホームページに係わる文書が同一の意味、内容をもつ保証はない。そこで、両言語のホームページに係わる文書が同一の意味、内容をもつか否かを判定し、同一の意味、内容をもつものであれば、両ホームページに記載される文章は対の関係にあり、非常に好ましい対訳データとなり、ひいては対訳データの量は大量に確保可能となる。

【0034】以上のようなインターネットのホームページの性質を踏まえて本発明システムを実現したものであり、以下、システムの動作を説明する。入力部11からの指示に基づき、表示部12にハイパーテキストの初期画面或いは所要の画面を表示した後、画面上の行き先（URL）をクリックすると、辞書構築支援処理部15は、辞書構築支援用記録媒体13に格納されるプログラムを読み取って辞書構築支援処理を実行する。

【0035】この辞書構築支援処理は、入力部11からの指定に基づく行き先（URL）アドレスおよびホームページ取り込み用コマンドを通信管理制御部17を介して例えばWWWサーバ21に発信し、当該WWWサーバ21の例えば英語ホームページに係わる文書を取り込んでデータ記憶部14に記憶する一方、表示部12に表示する。

【0036】この英語ホームページに係わる文書から例えば「日本語はここ」と記載されているとき、そのリンク先部分をクリックし、日本語のホームページも取り込んでデータ記憶部14に記憶する一方、表示部12に表示する。逆に、日本語ホームページに係わる文書を取り込み、その文章の中に例えば「for English」と記載されている場合も同様にそのリンク先部分をクリックし、英語ホームページに係わる文書を取り込む。

【0037】しかる後、辞書構築支援処理部15は、両言語のホームページに係わる文書の意味、内容の同一性を同一内容判定部16に依頼する。この同一内容判定部16は、両ホームページに係わる文書の中に存在するイメージデータ、グラフ、図柄、さらにリンクする先などが同一であるか否かを判定する。通常、同じイメージデータ、グラフなどが使用されているとき、両言語のホームページに記載される文章の意味、内容が同じである場合が多いためである。

【0038】さらに、必要に応じて、イメージデータ、グラフ、図柄、さらにリンク先などのデータが同一であることを条件に、或いはこれらイメージデータ、グラフ、図柄、リンク先などのデータの同一を条件とせず

に、両言語のホームページに記載される文章について、各言語の文法に基づいて文章を比較し、何%同一であるかを調べる。ここで、少なくとも80%以上同一であれば、他の残りの20未満の文章も同一であると判定し、辞書構築支援処理部15に通知する。

【0039】ここで、辞書構築支援処理部15は、日本語と英語のホームページに係わる文書から順次原文の単語と訳文の単語とを対の形で取り出し、対訳語としてデータ記憶部14または図示されていないが翻訳処理構成部分20の翻訳辞書に登録していく。

【0040】従って、以上のような実施の形態によれば、同一のWWWサーバの異なる両言語のホームページに係わる文書を取り込み、一定の条件のもとに同一性を調べ、両ホームページが同一であると判定するので、両言語のホームページに記載される文章を対訳データとして取得でき、国内および外国から発信されるホームページが無数存在することを考えれば、大量の対訳データを入手できる。しかも、辞書構築支援プログラムに基づいて自動的に辞書構築支援処理を実行するので、それ程コストおよび時間を必要とせずに対訳データを入手でき、また精度の翻訳辞書を容易に構築することができる。

(翻訳処理方法および翻訳処理用記録媒体に係わる発明の実施の形態) 図2は翻訳処理用記録媒体を用いた辞書構築支援システムの一実施形態を示す図、図3は翻訳処理用記録媒体のプログラムを用いて翻訳処理を行う翻訳処理方法を説明するフローチャートである。なお、図2において図1と同一機能部分には同一の符号を付して説明する。

【0041】この辞書構築支援システム10は、多数のWWWサーバ2、…をもつコンピュータネットワーク3に接続されている。入力部11は、ハイパーテキストの文書画面に基づきマウスなどで行き先(ULR)をクリックし、行き先となるWWWサーバの例えば日本語および日本語以外の言語のホームページに係わる文書を取り込む指示を行う。通常、キーボード、OCR、タブレット、マウス、ライトペン或いはFDD、ネットワークなどが用いられる。表示部12は、ハイパーテキストの文書画面の表示、WWWサーバから取り込むホームページに係わる文書の表示その他入力部11から入力されるデータやデータ記憶部14に記憶されるデータ等を表示するものであって、CRT、ビットマップディスプレイ、投影される壁面やゴーグルなどが用いられる。辞書構築支援用記録媒体13は図3に示すフローチャートに示すプログラムが格納されている。データ記憶部14は図1で説明したような必要なデータを記憶する。辞書構築支援処理本体部18は、CPUで構成され、辞書構築支援用記録媒体13に格納されるプログラムを読み取って、図1に示す辞書構築支援処理部15および同一内容判定部16と同様な機能を行う。機能分けすると、辞書構築

支援処理本体部18は、あるWWWサーバ2から両ホームページの文書を取り込むホームページ取り込み機能、これら両ホームページの同一性を判定する同一性判定機能、両ホームページの文書が同一であるとき、原文文書と対訳文書とを対の形で翻訳辞書に登録する対訳データ登録機能を実現する。

【0042】次に、辞書構築支援処理本体部18が辞書構築支援用記録媒体13に格納されているプログラムに基づいて図3に示す辞書構築支援処理を実行する。入力部11からの指示に基づき、表示部12にハイパーテキスト画面を表示した後、行き先(ULR)をクリックすると、CPUで構成される辞書構築支援処理本体部18は、辞書構築支援用記録媒体13のプログラムを読み取って辞書構築支援処理を実行する。

【0043】この辞書構築支援処理本体部18は、ホームページ取り込み機能を実行する。つまり、行き先のクリックによって指定されたWWWサーバ2のある言語で記載されるホームページAを取り込んでデータ記憶部14に記憶するとともに、表示部12に表示する(ST1)。しかる後、このホームページAの中に他の言語で書かれたホームページを示す部分があるかを判断する(ST2)。ここで、他の言語で書かれたホームページを示す部分があれば、そのホームページBを読み込んで同様にデータ記憶部14に記憶するとともに、表示部12に表示する(ST3)。

【0044】次に、辞書構築支援処理本体部18は同一性判定機能を実行する。この同一性判定機能は、両ホームページA、Bの中からイメージデータ、グラフ、図柄、さらにリンクする先などの情報C、Dを抽出する(ST4)。そして、これら情報C、Dが予め定めた割合(例えば80%)以上一致するか否かを判断する(ST5)。ここで、情報C、Dが予め定めた割合以上一致すると、両ホームページに記載される文章は同一であると判定する。

【0045】さらに、必要に応じて、イメージデータ、グラフ、図柄、さらにリンクする先などのデータが予め定めた割合以上一致することを条件に、或いはこれらイメージデータ、グラフ、図柄、リンクする先などのデータの一定割合以上の同一を条件とせずに、両ホームページに記載される文章について、各言語の文法に基づいて文章を比較し、どの程度の割合で同一であるかを調べる。ここで、少なくとも80%以上同一であれば、他の残りの20未満の文章も同一であると判定する。

【0046】引き続き、辞書構築支援処理本体部18は、両ホームページに記載される文章が同一であると判定されたとき、両ホームページから順次原文の単語と訳文の単語とを対の形で取り出し、対訳データとしてデータ記憶部14または機械翻訳処理システムの翻訳辞書に順次登録する(ST6)。

【0047】以上のような翻訳処理方法の実施の形態に

よれば、特別に専門辞書や対訳データを見つけ出すことなく、異なる両言語のホームページから大量の対訳データを迅速、かつ、低コストで適当な時間に入手しつつ、正確な翻訳辞書を容易に構築することができる。

【0048】また、以上のような翻訳処理用記録媒体を用いた辞書構築支援システムによれば、同一のWWWサーバのある言語のホームページから他の言語のホームページを容易に探し出て取り込むことができる。また、両ホームページの中から一定の条件のもとに両ホームページに記載される文章の同一性をほぼ正確に判定できる。さらに、両ホームページに記載される文章が同一と判定したとき、かかる文章を原文文書と対訳文書とを対の形で取り出し、対訳データとして翻訳辞書に登録できる。

【0049】なお、記録媒体としては、CD-ROMを用いるが、それ以外には例えば磁気テープ、DVD-ROM、フロッピーディスク、MO、MD、CD-R、メモ리카ードなどを用いてもよい。

(翻訳処理システムに係わる発明の実施の形態) 図4は本発明に係わる翻訳システムの一実施の形態を示す機能ブロック図である。

【0050】このシステムは、図1と同様に、WWWブラウザで作成されるハイパーテキストなどの文書をもつコンピュータによって構成される翻訳システム1がコンピュータネットワーク3を介して日本語以外の言語例えば第1の英語ホームページP1、そのリンク先となる第2の英語ホームページP2および日本語ホームページP3などをもつ少なくとも1台のWWWサーバ2、…が接続されている。

【0051】この翻訳システム1は、入力部31、表示部32、翻訳処理用記録媒体33、データ記憶部34、翻訳辞書や翻訳処理用文法を記憶する辞書・文法記憶部35、翻訳処理制御部36、訳語選択参照部37および通信制御部38等によって構成されている。

【0052】前記入力部31は、ハイパーテキストの文書画面表示指示その他必要なデータ、コマンドを入力するもので、通常、キーボード、OCR、タブレット、マウス、ライトペン或いはFDD、ネットワークなどが用いられ、ハイパーテキストの文書画面に基づきマウスなどで行き先(ULR)をクリックし、WWWサーバ2から例えば日本語および日本語以外の言語のホームページに係わる文書を取り込む機能をもっている。

【0053】前記表示部32は、ハイパーテキストの文書画面の表示、WWWサーバ2から取り込むホームページデータに係わる文書の表示その他入力部31から入力されるデータなどを表示するものであって、CRT、ビットマップディスプレイ、投影される壁面やゴーグルなどが用いられる。

【0054】前記翻訳処理用記録媒体33は、ある言語の原文を所要の言語に翻訳する時に用いられるプログラ

ムを格納するものであるが、その他例えば通信制御用プログラムや表示処理用プログラムなども格納されている。データ記憶部34は、ハイパーテキストの文書の表示画面データ、WWWサーバ2から取り込むホームページに係わる文書、入力部31から入力される翻訳原文その他翻訳処理によって得られる翻訳結果などを記憶する。辞書・文法記憶部35は、第1言語の原文単語を所要とする第2言語の単語に翻訳する訳語の他、第1言語の原文単語を第2言語の単語に翻訳するときの文法データが記憶されている。

【0055】これら翻訳処理用記録媒体33、データ記憶部34および辞書・文法記憶部35は、通常、磁氣的、電氣的、光學的、機械的などの記憶素子、つまり、FDD、CD-ROM、磁気ディスク、光ディスクなどが用いられており、かつ、これらが一体または別体の記録媒体であってもよい。

【0056】前記翻訳処理制御部36は、翻訳処理用記録媒体33に記憶される翻訳処理用プログラムを用いて翻訳処理を実行したり、入力部31から入力されるホームページデータの取り込みコマンドなどを出力したり、各構成部分を統括制御する機能をもっている。

【0057】前記訳語選択参照部37は、翻訳処理制御部36において例えば英語ホームページに係わる文書を翻訳する時、辞書・文法記憶部35から複数の訳語が抽出された時、英語ホームページに係わる文書のリンク先に日本語のホームページに係わる文書が存在するとは、この日本語のホームページに係わる文書中に含まれている語句を訳語として優先的に選択し、翻訳処理制御部36に提供する機能をもっている。

【0058】前記通信制御部38は、ULRアドレスの発信やホームページの取り込みコマンドの発信、ホームページデータの取り込みなどを行う機能をもっている。次に、以上のような構成の翻訳システムの動作について説明する。

【0059】入力部31から翻訳すべき原文を入力すると、翻訳処理制御部36は、入力部31からの原文をデータ記憶部34に記憶した後、当該原文に関し、辞書・文法記憶部35の辞書から所要の言語に関連する文法規則にしたがって訳語を取り出して表示部32に表示し、複数の訳語があるとき適宜選択し、翻訳結果としてデータ記憶部34に記憶する。そして、原文の翻訳結果を表示部32に表示する。

【0060】ここで、翻訳処理制御部36は、例えば原文の収集がハイパーテキストに基づくホームページに係わる文書の翻訳の場合には次のような処理を行う。まず、予め翻訳処理制御部36が通信制御部38を介してWWWサーバ2から例えば英語ホームページP1に係わる文書または当該文書のリンク先である例えば英語ホームページP2に係わる文書を取り込み、データ記憶部34に記憶し、かつ、表示部32に表示する。そして、翻訳処

理制御部6において取り込んだ文書について、辞書・文法記憶部35の辞書および文法に従って翻訳処理を行う。

【0061】ところで、この原文である英語のホームページに係わる文書のある単語を翻訳する際、辞書・文法記憶部35から複数の訳語が抽出されたとき、前記WWWサーバ2の英語ホームページに係わる文書P1またはこの文書P1のリンク先である英語ホームページに係わる文書P2にリンクされている訳語言語つまり日本語ホームページに係わる文書P3があれば、当該文書P3を取り込んで訳語選択参照部37に記憶する。そして、この日本語ホームページに係わる文書P3の中に文書P1またはP2の単語に対応する訳語があれば、その訳語を優先的に選択し翻訳結果とするので、翻訳品質の向上を図ることができる。

【0062】従って、以上のような実施の形態によれば、WWWサーバ2の原文の文書を翻訳するとき、第1義的には辞書・文法記憶部35の訳語を用いて翻訳するが、複数の訳語あるとき、WWWサーバ2の原文文書にリンクされている訳語言語の文書の中に対応する訳語があるとき、その訳語を優先的に選択し翻訳結果とするので、翻訳品質の向上を図ることができる。

(翻訳処理方法および翻訳処理用記録媒体に係わる発明の実施の形態) 図5は翻訳処理用記録媒体の翻訳処理用プログラムを読み取って所定の機能を実現する翻訳処理システムの一実施形態を示す図、図6は翻訳処理用プログラムを用いて翻訳処理を行う翻訳処理方法を説明するフローチャートである。なお、図5において図4と同一機能部分は同一の符号を付して説明する。

【0063】この翻訳処理システム1は、多数のWWWサーバ2、…をもつコンピュータネットワーク3に接続されている。入力部11(通信制御部38)は、翻訳しようとする原文を入力する部分であって、例えばハイパーテキストの文書画面に基づきマウスなどで行き先(URL)をクリックし、行き先となるWWWサーバから例えば原文である英語ホームページに係わる文書または当該文書のリンク先である英語ホームページに係わる文書を取り込む指示を入力し、さらに翻訳処理制御指示や前記ホームページに係わる文書と同一の意味、内容をもつ日本語ホームページに係わる文書を取り込む指示を入力するものであって、通常、キーボード、OCR、タブレット、マウス、ライトペン或いはFDD、ネットワークなどが用いられる。表示部12は、ハイパーテキストの文書画面の表示、WWWサーバから取り込むホームページデータの表示、その他入力部11から入力される原文や翻訳結果を表示するものであって、CRT、ビットマップディスプレイ、投影される壁面やゴーグルなどが用いられる。

【0064】前記翻訳処理用記録媒体33は、ある言語の原文を所要の言語に翻訳する時に用いられるプログラ

ムを格納するものであって、例えば図6に示すようなフローチャートに示す翻訳処理を実行する。データ記憶部34は、ハイパーテキストの文書の表示画面データ、WWWサーバ2から取り込む原文となるホームページに係わる文書またはそのリンク先の文書、その他翻訳結果のデータを記憶するものである。辞書・文法記憶部35は、第1言語の原文単語を所要とする第2言語の単語に翻訳する訳語を記憶する他、第1言語の原文単語を第2言語の単語に翻訳するときの文法データが記憶されている。これら翻訳処理用記録媒体33、データ記憶部34および辞書・文法記憶部35は、通常、磁氣的、電氣的、光学的、機械的などの記憶素子、つまり、FDD、CDROM、磁気ディスク、光ディスクなどが用いられており、かつ、これらが一体または別体の記録媒体であってよい。

【0065】前記翻訳処理制御本体部39は、CPUで構成され、翻訳処理用記録媒体13に格納されている翻訳処理プログラムを読み取って図6に示す翻訳処理を実行する機能をもっている。この翻訳処理制御本体部39について機能分けすると、WWWサーバ2から原文となる日本語以外の言語例えば英語ホームページに係わる文書およびその文書のリンク先の英語ホームページに係わる文書を取り込む原文取り込み機能、この原文取り込み機能によつて取り込んだ原文に基づいて翻訳辞書から訳語を取り出して翻訳する翻訳処理機能、この翻訳処理機能の段階で複数の訳語を抽出したとき、前記英語ホームページに係わる文書からリンク先の日本語ホームページに係わる文書を取り込んで参照し、この文書に含まれる語句を優先的に訳語として選択する訳語選択参照機能を実現する。

【0066】次に、翻訳処理制御本体部39が翻訳処理用記録媒体33に格納されるプログラムに基づいて図6に示す翻訳処理を実行する。翻訳処理制御本体部39においては、翻訳処理用記録媒体33のプログラムに基づいて原文取り込み機能を実現する。

【0067】すなわち、入力部31からの指示に基づき、ハイパーテキストの文書に係わる翻訳か否かを判断する(ST11)。ここで、一般の翻訳処理の場合には、従来公知の手段、つまり入力部31から入力される翻訳原文を受け、辞書・文法記憶部35の訳語を抽出し選択確認して翻訳結果を得る(ST12)。

【0068】ステップST11においてハイパーテキストの文書に係わる翻訳である場合には、行き先(URL)リンクをクリックするとともに、ホームページ取り込みコマンドを発信し、WWWサーバ2から原文となる例えば英語ホームページに係わる文書およびその文書のリンク先の英語ホームページに係わる文書を取り込み、データ記憶部34に記憶するとともに、表示部32に表示する(ST13)。

【0069】しかる後、翻訳処理制御本体部39は、翻

訳処理用記録媒体 33 のプログラムに基づいて翻訳処理を実現する。この翻訳処理は、原文となる例えば英語ホームページに係わる文書について辞書・文法記憶部 35 の文法に従って辞書から順次訳語を抽出した後、原文である文書のある単語に関し、複数の訳語があるか否かを判断する (ST14)。ここで、訳語が 1 つしかないとき、その訳語を用いて翻訳結果を得る。

【0070】次に、翻訳処理制御本体部 39 は、翻訳処理用記録媒体 33 のプログラムに基づいて訳語選択参照機能を実現する。この訳語選択参照機能は、ステップ ST14 においてある原文の 1 つの単語に複数の訳語が存在する場合、前記 WWW サーバ 2 の前記英語ホームページに係わる文書のリンク先に日本語ホームページに係わる文書があるか否かを判断し (ST15)、日本語ホームページに係わる文書があれば、その文書を取り込んで訳語選択参照部 35 を格納する (ST16)。そして、前記原文である英語ホームページに係わる文書と日本語ホームページに係わる文書とを比較参照し、当該日本語ホームページに係わる文書の中に原文単語に対応する日本語の語句があれば、それを優先的に選択し、原文単語の訳語として採用し翻訳処理を実行する (ST17)。

【0071】従って、以上のような翻訳処理方法の実施の形態によれば、WWW サーバ 2 のハイパーテキストの文書である例えば英語ホームページに係わる文書またはそのリンク先である文書を翻訳するとき、一義的には辞書・文法記憶部 35 を用いて翻訳するが、その原文である文書のある単語に関し、複数の訳語が存在するとき、英語ホームページに係わる文書のリンク先に日本語ホームページに係わる文書または日本語に係わる文書があるとき、その日本語ホームページに係わる文書または日本語に係わる文書を参照し、その中から原文である文書のある単語に対応する日本語の訳語を優先的に選択し翻訳するので、辞書・文法記憶部 35 に選択訳語が不足している場合でも確実に翻訳できる。また、ハイパーテキストの文書においてその文書内にリンク先として英語に係わる文書および日本語に係わる文書があるとき、これらは内容的に同じ分野の文書で記述していることが多いので、日本語の訳語を優先的に選択して翻訳すれば、翻訳品質の向上を図ることができ、さらに翻訳処理の効率を上げることができる。

【0072】また、以上のような翻訳処理用記録媒体 33 を用いた翻訳処理システムによれば、WWW サーバ 2 から原文となる例えば英語ホームページに係わる文書およびその文書のリンク先の英語ホームページに係わる文書を取り込み、辞書・文法記憶部 35 の文法に従って辞書から順次訳語を抽出し翻訳結果を得ることができる。

【0073】また、原文の 1 つの単語に複数の訳語が存在する場合、前記 WWW サーバ 2 の英語ホームページに係わる文書のリンク先に日本語ホームページに係わる文書を取り出し、その文書を比較参照しながら最適な語句

を見つけ出して翻訳処理を実現できる。

【0074】

【発明の効果】以上説明したように本発明によれば、次のような種々の効果を奏する。請求項 1~4 の発明では、WWW サーバから異なる言語のホームページに係わる文書を取り込み、両言語のホームページに係わる文書が同一性を有するとき、両言語のホームページに係わる文書から原文と訳語との対をなす訳文データを入手するので、大量の訳文データを入手可能であり、しかも低コスト、かつ、迅速に翻訳辞書を構築可能である。

【0075】請求項 5 の発明では、WWW サーバから取り込んだ異なる言語のホームページに係わる文書に基づいて翻訳辞書の構築策用対応文書を容易に取得可能な辞書構築支援用記録媒体を提供できる。

【0076】請求項 6 および 7 の発明では、WWW サーバから原文となるホームページに係わる文書を取り込んで翻訳するに際し、複数の訳語が存在する時、原文となるホームページに係わる文書のリンク先に訳語となるホームページに係わる文書があれば、当該文書に存在する訳語を優先的に選択し翻訳するので、適切な訳語を選択して精度の高い翻訳を行うことができる。

【0077】請求項 8 の発明では、WWW サーバから原文となるホームページに係わる文書を取り込んで翻訳するに際し、複数の訳語が存在する時、ホームページに係わる文書にとって適切な訳語を選択できる翻訳処理用記録媒体を提供できる。

【図面の簡単な説明】

【図 1】 本発明に係わる辞書構築支援システムの一実施の形態を示す構成図。

【図 2】 辞書構築支援用記録媒体を説明するためのシステム構成図。

【図 3】 図 2 に示すシステムの動作および辞書構築支援方法を説明するフローチャート。

【図 4】 本発明に係わる翻訳システムの一実施の形態を示す構成図。

【図 5】 翻訳処理用記録媒体を説明するためのシステム構成図。

【図 6】 図 4 に示すシステムの動作および翻訳処理方法を説明するフローチャート。

【符号の説明】

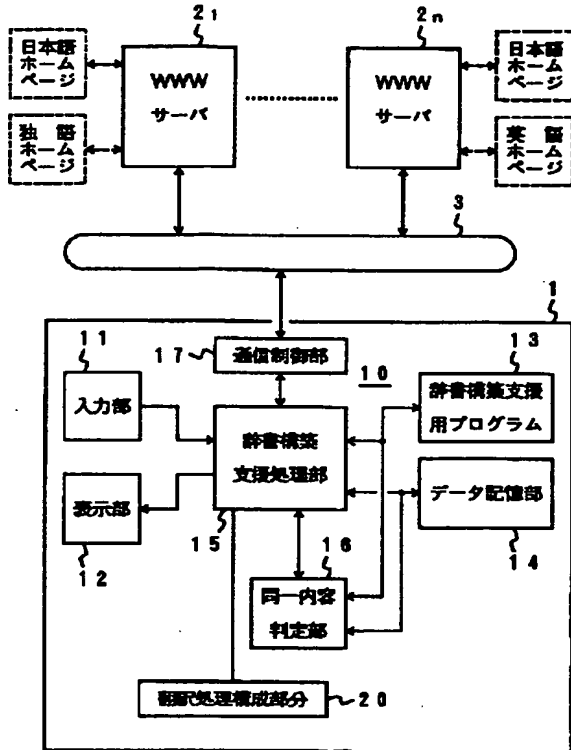
- 1…翻訳システム
- 2, 2₁, …, 2_n…WWWサーバ
- 3…コンピュータネットワーク
- 10…辞書構築支援システム
- 13…辞書構築支援用記録媒体
- 15…辞書構築支援処理部
- 16…同一内容判定部
- 18…辞書構築支援処理本体部
- 20…翻訳処理構成部分
- 21…ユーザ端末機

17

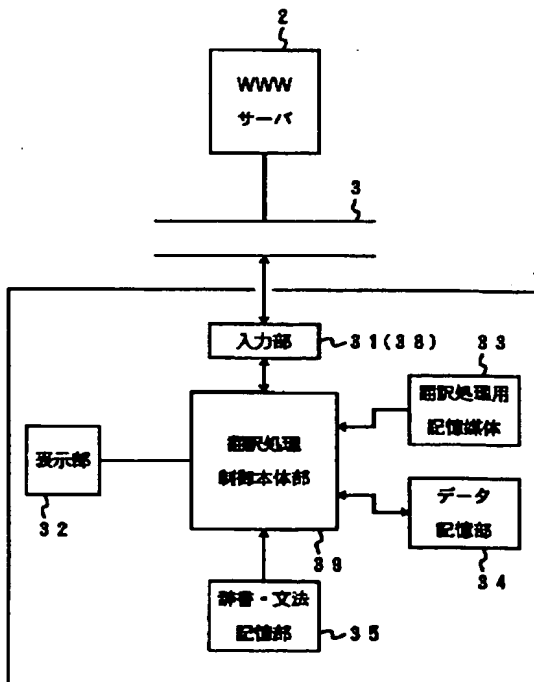
33...翻訳処理用記録媒体

36...翻訳処理制御部

【図1】



【図5】

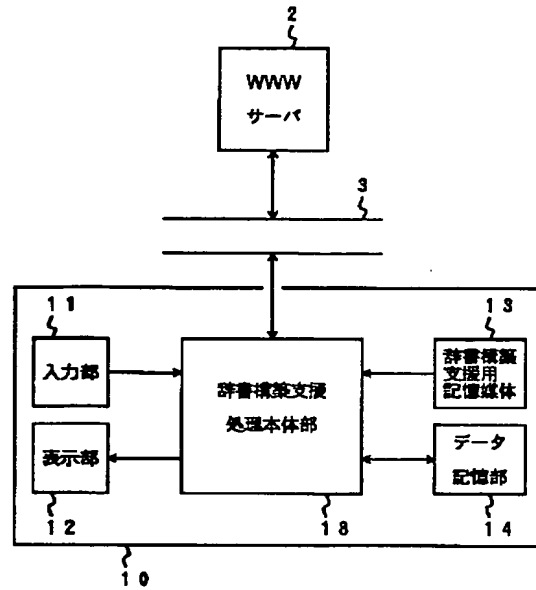


18

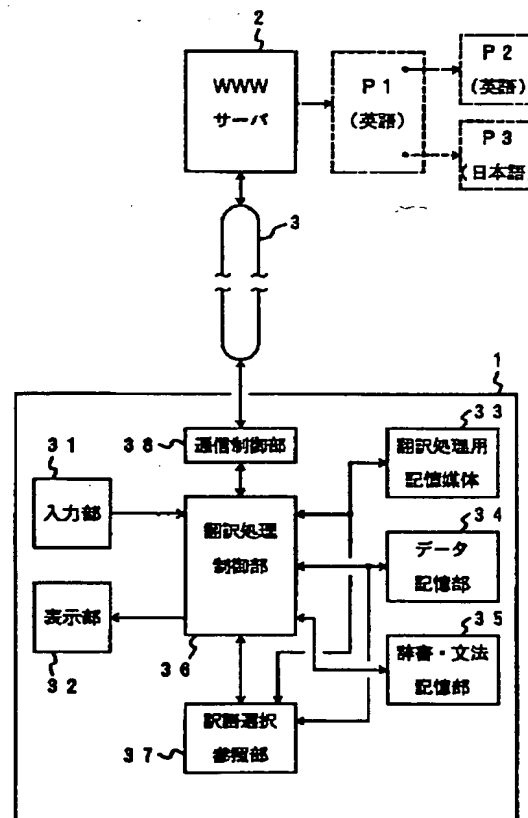
37...訳語選択参照部

39...翻訳処理制御本体部

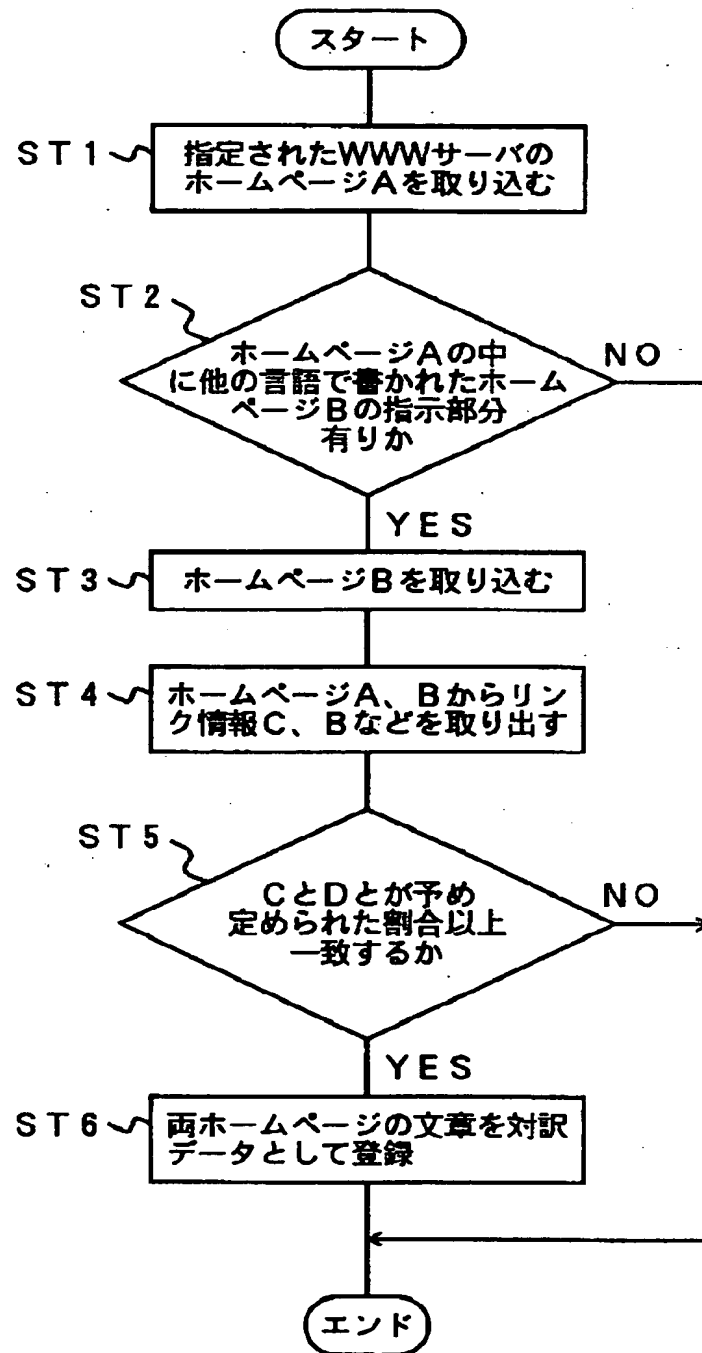
【図2】



【図4】



【図3】



【図6】

